A coliform-targeted metagenomic method facilitating human exposure estimates to *Escherichia coli*-borne antibiotic resistance genes

Authors
Leonard, A. F. C.[a], Yin, X. L.[b], Zhang, T.[b], Hui, M[c]., Gaze, W. H[a1].

Affiliations
a) European Centre for environment and Human Health, University of Exeter Medical School, Knowledge Spa, Royal Cornwall Hospital, Truro, United Kingdom TR1 3HD
b) Department of Civil Engineering, The University of Hong Kong, Hong Kong
c) Department of Microbiology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong

Key words:
Antibiotic resistance; *Escherichia coli;* Bathing waters; Recreation; High-throughput sequencing

Abstract

Background: Antimicrobial resistance and the spread of antibiotic resistance genes (ARGs) poses a threat to human health. Community-acquired infections resistant to treatment with first-line antibiotics are increasing, and there are few studies investigating environmental exposures and transmission.

Aim: To develop a novel targeted metagenomic method to quantify the abundance and diversity of ARGs in a faecal indicator bacterium, and to estimate human exposure to resistant bacteria in a natural environment.

Approach: Sequence data from *Escherichia coli* metagenomes from 13 bathing waters in England were analysed using the ARGs Online Analysis Pipeline to estimate the abundance and diversity of resistance determinants borne by this indicator bacterium. These data were averaged over the 13 sites and used along with data on the levels of *E. coli* in English bathing waters in 2016 and estimates of the volume of water that water users typically ingest in an average session of their chosen activity, to quantify the numbers of ARGs that water users ingest.

Findings: *E. coli* in coastal bathing waters were found to harbour on average 1.24 ARGs per cell. Approximately 2.5 million water sports sessions occurred in England in 2016 which resulted in water users ingesting at least 100 *E. coli*-borne ARGs.

[1] Corresponding author: European Centre for Environment and Human Health, University of Exeter Medical School, Knowledge Spa, Royal Cornwall Hospital, Truro, UK TR1 3HD; Tel +44 (0) 1872 258149; Email w.h.gaze@exeter.ac.uk

Introduction

Antimicrobial resistance is one of the greatest public health threats faced today. Infections caused by resistant organisms are on the rise, with resulting treatment failure and increases in morbidity, mortality, and costs associated with increased demands on healthcare services and loss in productivity. According to a recent review of antimicrobial resistance, if current trends continue, 10 million deaths are predicted to occur annually by the year 2050, with a global economic cost of $100 billion USD by this time (Review on antimicrobial resistance 2014).

Much research concerning the transmission and acquisition of resistant infections has focused on the role of hospitals. However, the frequency of resistance to antibiotics among community-acquired pathogens is increasing, and with up to 80% of antibiotics prescribed in primary care settings, research into non-nosocomial transmission routes is of increasing importance (Costelloe, et al. 2014). One way in which people might be exposed to and infected by resistant bacteria is through exposure to natural waters contaminated by faecal pollution. Natural aquatic environments, such as rivers and oceans, are regularly contaminated by faecal pollution from anthropogenic and zoonotic sources, which introduces large numbers of microorganisms to these aquatic environments, some of which can be pathogenic and/or harbour clinically important antibiotic resistance genes (ARGs) (Blackburn, et al. 2017, Pruden, et al. 2006, Review on antimicrobial resistance 2015). Approximately 50% of the world's population lives within 100 km of the coast (Shuval 2003), and in the UK, millions of people visit the coast each year (Kantar TNS 2017), and engage in activities involving head immersion in seawater and water ingestion (Leonard, et al. 2015). Very few studies have investigated these natural environments as sites where people could be exposed to and infected by resistant bacteria. The few that have investigated transmission of resistant bacteria via aquatic environments used culture-based methods to isolate bacteria phenotypically resistant to antibiotics, and found evidence of an association between exposure to natural waters and colonisation by resistant bacteria (Leonard, 2018) or community-acquired urinary tract infections caused by resistant bacteria (Soraas, et al. 2013).

High-throughput sequencing methods offer opportunities to explore the diversity and abundance of ARGs in environmental samples and reduce biases introduced by PCR-based methods and cultivation of phenotypically resistant bacteria (Li, et al. 2015). These techniques are increasingly used to study environmental microbiology and the ecology of resistant bacteria in various matrices, and are able to produce data on unculturable bacteria (Ercolini 2013, Logares, et al. 2012). However a problem often identified with culture-independent sequencing of bacteria is a lack of information on the bacterial hosts of identified genes in samples, many of which may be harmless to humans.

The primary aim of this study was to develop a novel approach using targeted metagenomics to investigate the relative abundance and diversity of antibiotic resistance genes (ARGs) in the faecal indicator bacterium, *Escherichia coli*, in coastal bathing waters. *E. coli* were selected as indicator organisms because of their presence in the gut microbiomes of animals (including humans), and their ease of culture leading to widespread use to monitor water quality (including drinking water, bathing waters, and shellfish waters) (International Organization for Standardization 2005, Monaghan and Hutchinson 2010, Odonkor and Ampofo 2013). In addition *E. coli* also readily undergo horizontal gene transfer with other members of the microbial community (Handel, et al. 2015, Levy, et al. 1976) and clinically important antibiotic resistance genes can easily spread from non-pathogenic bacteria to ones capable of causing disease in humans (Ashbolt, et al. 2013). *E. coli*, for example *E. coli* sequence type 131 (Johnson, et al. 2010) and *E. coli* O157:H7 (Ihekweazu, et al. 2006), are increasingly important clinical pathogens causing a wide variety of intestinal and extra-intestinal infections (Russo and Johnson 2003). *E. coli* are the leading cause of bloodstream infections, responsible for causing 40,580 cases of bacteraemia in England between 2016 and 2017, more than three times greater than the number of bacteraemias caused by methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* combined (Public Health England, 2017). *E. coli* are therefore useful indicator organisms for monitoring antibiotic resistance in aquatic environments, such as bathing waters, to which humans have high exposure levels.

A secondary aim was to use data generated by this study to estimate bather exposure to ARGs borne by *E. coli* in coastal bathing waters in England in 2016.

Methods

Isolation of *Escherichia coli* from coastal bathing waters

The Environment Agency, as part of their routine monitoring of designated bathing waters (European Parliament Council of the European Union 2006), collect water samples from 415 bating waters in England on a weekly basis every year between mid-May and the end of September (Environment Agency 2017). The levels of *E. coli* in water samples are quantified using standard methods for membrane filtration (performed by Environment Agency staff), in which 0.1 mL, 1 mL, and 10 mL water samples are passed through a filter (0.45 µm pore size) and these filters are incubated on Tryptone Bile X-Glucuronide agar for 4±1 hours at $30^0C \pm 1^0C$, followed by 14±2 hours at $44^0C \pm 0.5^0C$ (Environment Agency 2009, International Organization for Standardization 2014a). After incubation, all blue colonies on the filter are counted as confirmed *E. coli*. Validation confirmed that 99% of phenotypic *E. coli* colonies are genotypic *E. coli* (unpublished data, Jonathan Porter, Environment Agency, pers. comm), meaning very few non-*E. coli* species will be included in the *E. coli* composite samples.

For this study, the Environment Agency provided filters with cultured *E. coli* from water samples taken at thirteen different designated coastal bathing waters. All water samples were collected between 21 September 2016 and 22 September 2016 and were transported on ice from the Environment Agency Laboratory (Starcross, Exeter) to the University of Exeter Medical School laboratory for processing (Penryn, Cornwall).

Sample preparation and DNA extraction

All *E. coli* colonies from each site were counted, and every *E. coli* colony that was growing separately (i.e. not touching other colonies) was picked using a sterile 1000 µL pipette tip and combined in an Eppendorf with 0.85% sterile NaCl solution to form a composite *E. coli* metagenome sample for each site. In addition, all the non-*E. coli* coliform (NEC) colonies were picked into a composite NEC sample across all 13 sites. DNA was extracted from the 14 composite samples using the FastDNA ™ Spin kit (MP Biomedicals) according to the manufacturer's instructions, followed by RNAse digestion and clean up (Agenourt AMPure XP System Beckman Coulter). DNA concentrations were determined using the Qubit ™ dsDNA high sensitivity assay kit (Invitrogen). The DNA was sent for sequencing (University of Exeter, Exeter Sequencing Service) which provided Nextera XT library preparation and Illumina high-throughput sequencing on MiSeq, generating 2 x 250 bp paired end reads.

The amount of biomass contributed to the composite samples by each *E. coli* colony was not standardised by cell count. Therefore, larger colonies could contribute more genetic material to a sample than smaller colonies, and genes in larger colonies be over-represented in the absolute abundance of genes in each sample. To assess whether the methods used to pick cells from filters resulted in similar numbers of cells contributing biomass to the composite samples, the number of *E. coli* picked using a 1000 µL pipette tip was quantified for 18 *E. coli* colonies (see supplementary material for methods).

Bioinformatic and statistical analyses

Sequence data underwent analysis using the validated Antibiotic Resistance Genes Online Analysis Pipeline (Yang, et al. 2016). Sample sequences were aligned against sequences in the Structured ARG (SARG) database which contains 4049 ARG sequences. Sample sequences were identified as ARGs based on identity being more than 85% and alignment length ratio being more than 75%. Results were categorised into resistance 'type' (24) and 'subtype' (1209) whereby 'type' is the class of antibiotic to which the ARG confers resistance (e.g. beta-lactams), and 'subtype' is the identity of the resistance gene (e.g. CTX-M-15 of beta-lactams). The abundance of ARGs was quantified along with the relative abundance of genes per *E. coli* cell, with *E. coli* abundance being calculated as previously

described (Yang, et al. 2016), and the Simpson Index of diversity of ARGs at each site were calculated using the R package, phyloseq.

Water user exposure to ARGs borne by *E. coli* in coastal bathing waters

The mean number of resistance genes per *E. coli* over all 13 sites was calculated for each type of antibiotic class and are presented in Table S2. However, for the main analysis, data on multi-drug resistance (MDR) genes (82 subtypes) and unclassified resistance gene (17 subtypes) were excluded because although they contribute to the antibiotic resistance potential, they may be present in both antibiotic-susceptible and antibiotic-sensitive bacteria (Li, et al. 2015). The mean number of genes for each antibiotic per *E. coli* were then summed to give the average number of resistance genes per *E. coli*.

Similar methods to those described by Leonard et al. (2015) were used to estimate the number of *E. coli*-borne ARGs ingested by coastal water users in England in 2016. Briefly, data on the density of *E. coli* in English designated bathing waters were obtained from the Environment Agency website for the 2016 bathing season (Environment Agency 2017). *E. coli* data from bathing water sample were categorised into three groups based upon *E. coli* density data: 1) *E. coli* density ≤250 *E. coli* colony-forming units (cfu) per 100 mL; 2) *E. coli* density between 251 and 500 cfu 100 mL^{-1}; 3) *E. coli* density >500 cfu 100mL^{-1}. *E. coli* densities were converted from cfu 100 mL^{-1} to cfu mL^{-1} and multiplied by the number of resistance genes carried by the average *E. coli*. These were then applied to the average volume of water that various water users ingest during a typical session of their chosen activity to give *X*, the number of resistance genes borne by *E. coli* ingested by water users:

$$X = E.R.V$$

Where *E* is the number of *E. coli* cfu mL^{-1}, *R* is the number of resistance genes carried by the average *E. coli*, and *V* is the volume of water ingested (mL).

The mean of *X* was then estimated for each water sport by water quality category.

The number of water sports sessions in which *E. coli*-borne ARGs were ingested by water users in England in 2016 was then estimated (Leonard, et al. 2015). The Office for National Statistics estimated population of England in mid-2016 to be 55,268,200, with 13,107,000 aged 19 or under, and 42,161,200 aged 20 and above (Office for National Statistics 2016). A variety of estimates were produced to understand the scale of exposure if different thresholds of resistance gene carriage were considered: exposure to at least one ARG borne by *E. coli*, at least 10 *E.coli*-borne ARGs, 100 *E.coli*-borne ARGs and 1000 *E.coli*-borne ARGs.

Results

Isolation of *E. coli* from coastal bathing waters

Filters with *E. coli* and other coliform colonies from 13 designated coastal bathing sites were acquired: six from beaches in the south of England and seven from beaches in the northeast (Table 1). A total of 315 *E. coli* colonies were picked across all thirteen sites (mean number of *E. coli* per site = 24, standard deviation 21). In addition, all 105 non-*E. coli* coliform (NEC) colonies from across the 13 sites were picked into a single composite NEC sample.

Site number	Region	Number of *E. coli* colonies sampled	Number of *non-E. coli* coliform colonies sampled
1	South	15	20
2	South	20	16
3	South	20	13
4	South	10	5
5	North east	40	3
6	South	12	2
7	North east	20	1
8	North east	18	1
9	North east	27	2
10	North east	90	5
11	South	21	5
12	North east	9	14
13	North east	13	18
Total		315	105

Table 1 The number of *E. coli* colonies and non-*E. coli* colonies picked at each site, and the region of England in which each site was located.

The numbers of *E. coli* picked by a 1000 µL pipette tip was in the same order of magnitude, and the majority of replicates were within one standard deviation of the mean (2.41×10^8) (Supplementary Figure S1). Using a pipette tip was a fast and simple method of picking consistent numbers of cells. Furthermore, the number of *E. coli* colonies picked at each site appears to have no association with the abundance of ARGs identified at each site, although given the small sample size, it is difficult to see relationships (Figure S2).

Results of sequence analyses

The average library size was 356MB and assuming a typical *E. coli* genome of 5MB (Lukjancenko, et al. 2010), this gave a coverage equivalent to >70 genomes per composite sample. The sequencing depth for each sample is reported in Table S1. Among all 13 sites, resistance genes to 12 different classes of antibiotic were identified in *E. coli*. All 13 sites had *E. coli* with resistance genes to at least seven different antibiotic classes, and one site had resistance genes to 11 classes of antibiotic (Figure 1). Resistance genes against 10 types of antibiotics were not detected. The mean number of resistance genes per *E. coli* across all sites was 1.24 (Table S2).

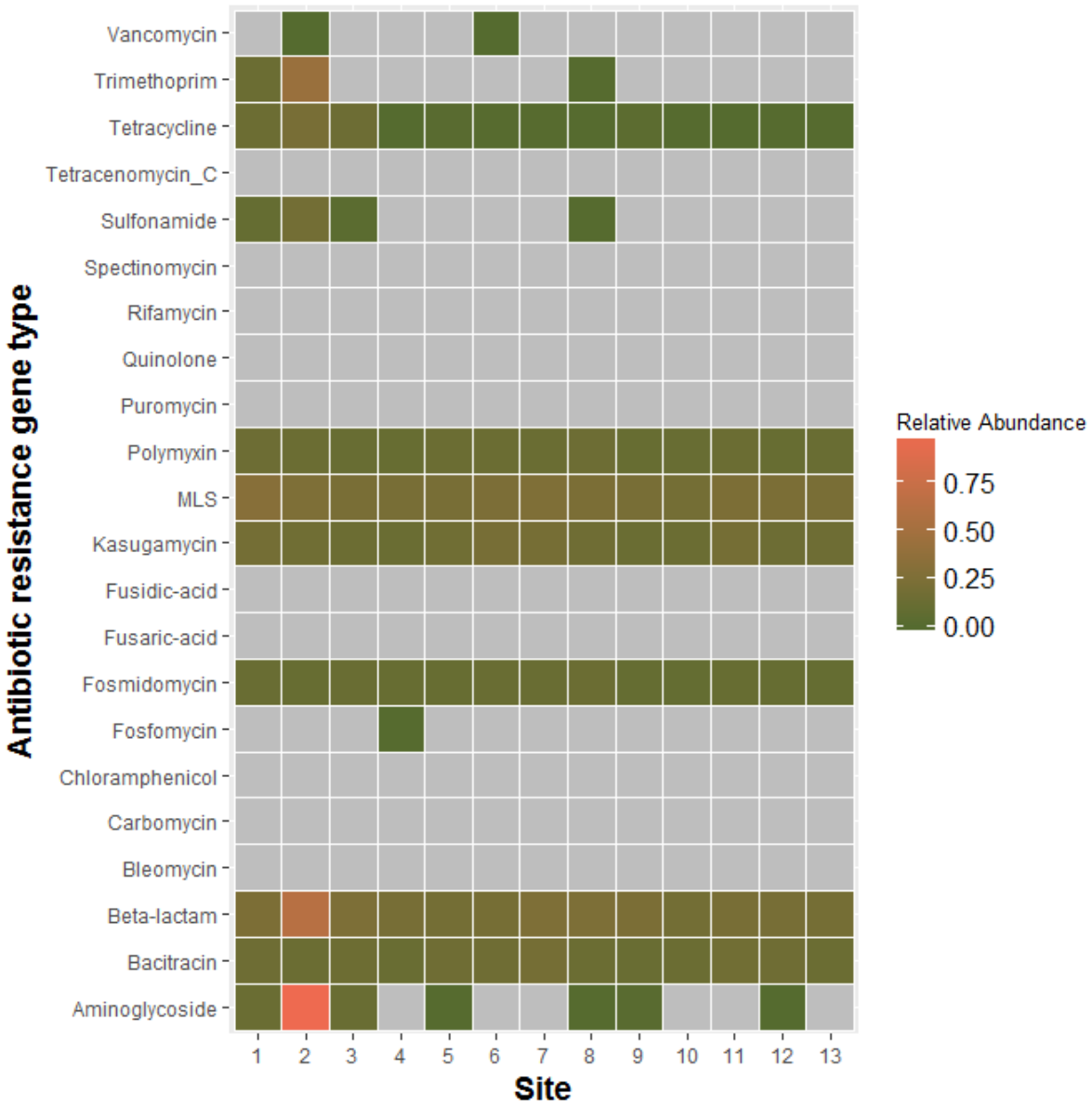

Figure 1 The abundance of resistance genes to types of antibiotic per *E. coli* at each sample site. High relative abundance is indicated by pink, low relative abundance is coloured in green. Grey cells indicate no detection. MLS is the antibiotic type macrolide-lincosamide-streptogramin.

The relative abundance of resistance determinants in NEC from all sites was also quantified. While there was a lower abundance of genes compared to *E. coli* (0.19 resistance genes per 16S rRNA in NEC compared to 0.45 per 16S rRNA in *E. coli*) (Table S2), genes for resistance to 13 different classes of antibiotic were detected in the NEC, some of which were not found in *E. coli* (e.g. chloramphenicol and quinolone). Genes conferring sulphonamide resistance were detected in *E. coli* but not in NEC (Figure S3).

Resistance genes against the beta-lactam class of antibiotics was the most common type found among *E. coli*, accounting for 20% of the resistance genes across all sites (range 14% to 24%). Similarly, ARGs against beta-lactams were the most prevalent type of resistance among NEC, making up 31% of the resistance genes detected. The percentage of the total numbers of genes that confer resistance to each class of antibiotic among *E. coli* and NEC are displayed in Figure 2.

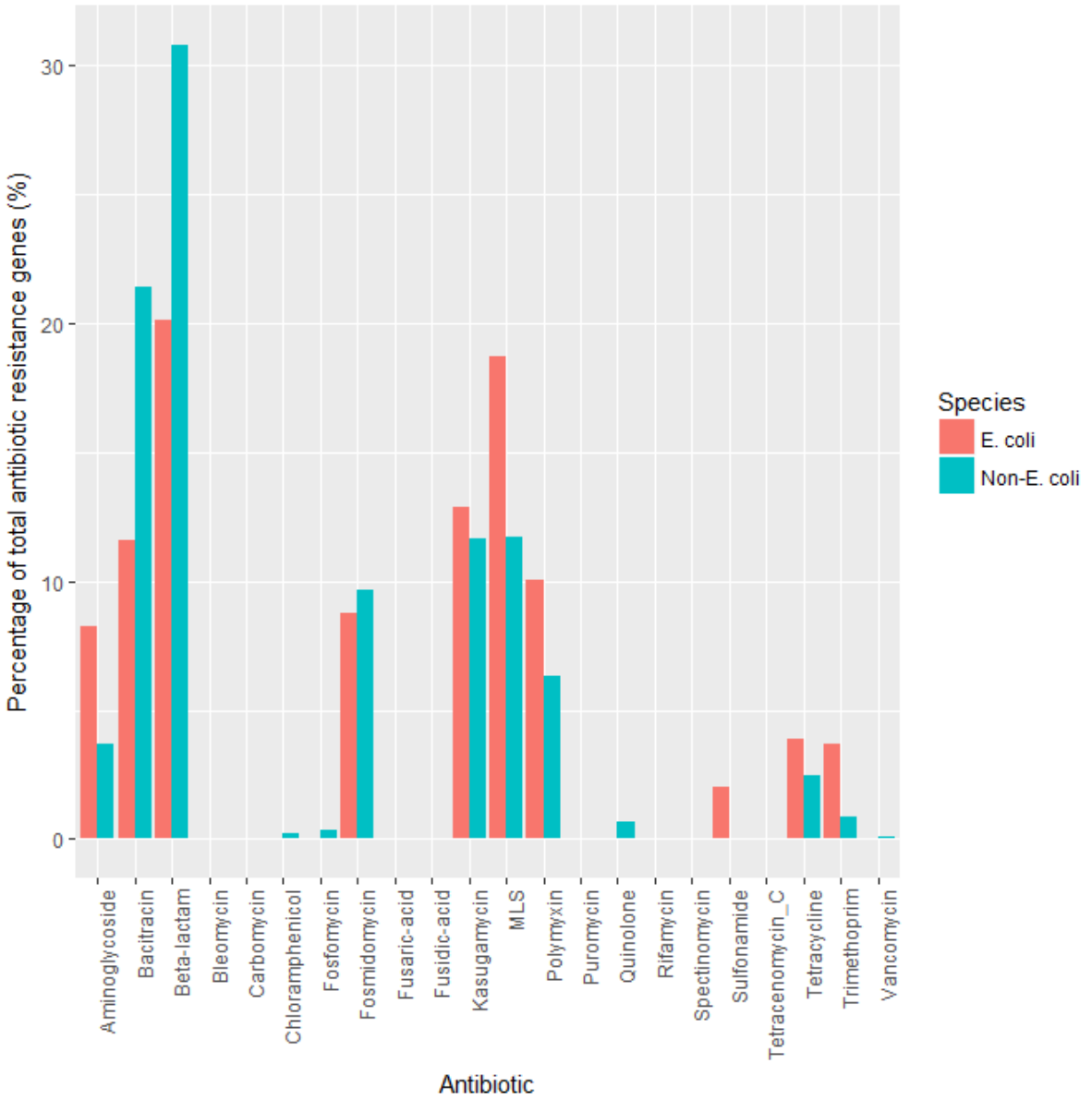

Figure 2 The percentage of genes conferring resistance to each type of antibiotic in the total *E. coli* and total non-*E. coli* coliforms populations. MLS is the antibiotic type macrolide-lincosamide-streptogramin.

The Simpson index of diversity at each site was quantified using data on the gene subtype count among *E. coli* at each site. *E. coli* from site number 2 had the highest Simpsons Index of diversity compared to the *E. coli* from other sampled sites (Figure 3), although samples from sites 1 and 3 also demonstrated high levels of ARG diversity compared to the 10 remaining samples. The diversity of ARGs among the composite NEC sample was also high. Other indices of diversity are displayed in Figure S4.

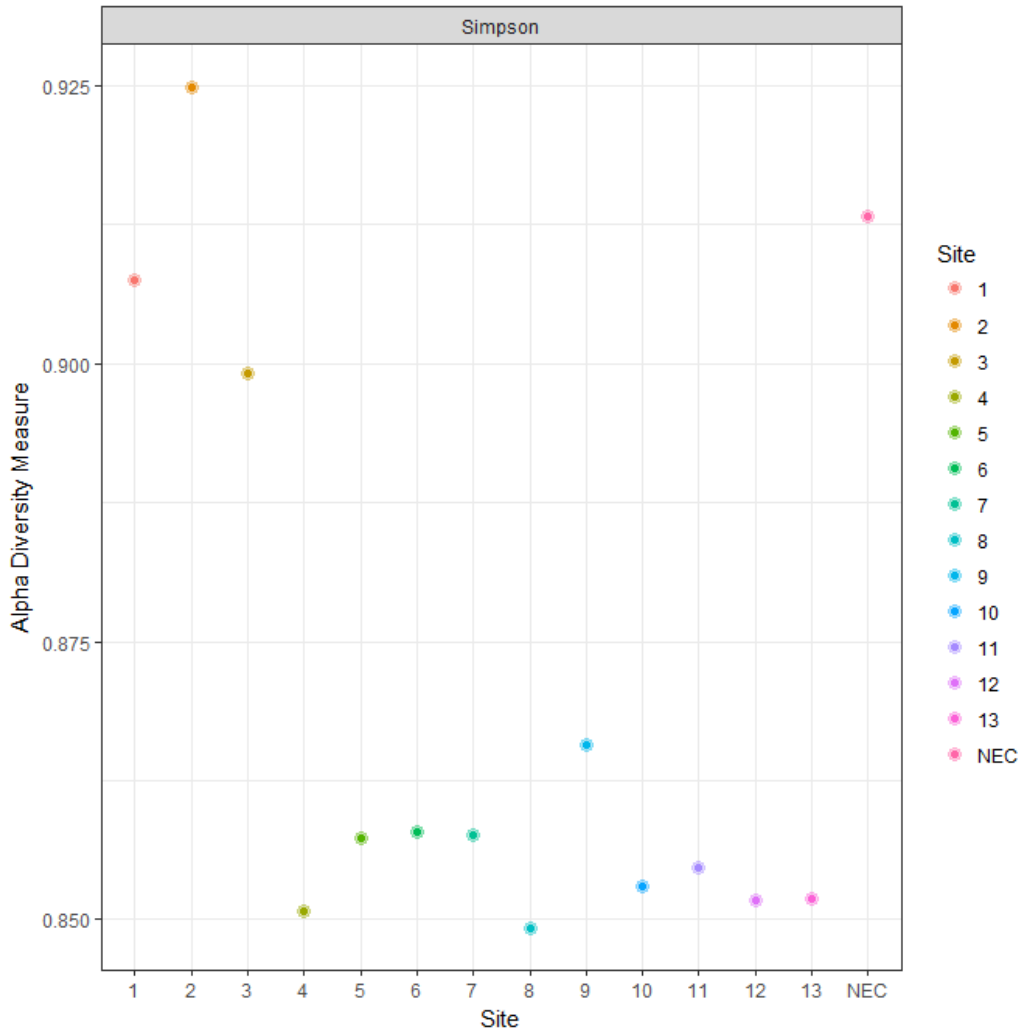

Figure 3 Simpson index of diversity of *E. coli*-borne antibiotic resistance genes at the 13 sample sites and among non-*E. coli* coliforms at all sites (NEC).

Bather exposure

The average *E. coli* in English bathing waters in 2016 was estimated to harbour 1.24 ARGs. Table 2 presents the mean number of *E. coli*-borne ARGs ingested during a typical session of various high- and low-contact water sports.

This data was used to calculate a population-level estimate of exposure to ARGs borne by *E. coli* in coastal bathing waters. Of the 8218 water samples collected by the Environment Agency between 15 May 2016 and 30 September 2016, 7874 (95.8%) had *E. coli* densities of ≤250 cfu 100 mL^{-1}. 205 (2.49%) had *E. coli* densities between 250 and 500 cfu 100 mL^{-1}, and 139 (1.69%) had *E. coli* densities >500 cfu 100 mL^{-1} (Supplementary Table S3). Over 123 million water sports sessions occurred in England in 2016 which resulted in the ingestion of at least one ARG borne by *E. coli* in coastal bathing waters (Table 3), representing 100% of water sports sessions involving ingestion of resistant *E. coli*. Nearly 2.5 million sessions occurred if the threshold of exposure is raised to the ingestion of *E. coli* harbouring at least 100 ARGs.

	Mean number of *E. coli*-borne antibiotic resistance genes ingested by different water users in bathing waters in different categories of water quality		
	E. coli density ≤250 cfu 100 mL^{-1}	*E. coli* density 251-500 cfu 100 mL^{-1}	*E. coli* density >500 cfu 100 mL^{-1}
High-contact water sports*			
Swimming (adults)	5.70	69.3	280
Swimming non-adults	13.2	160	648
Surfing	60.7	739	2987
Diving	3.52	42.9	173
Low-contact water sports*			
Boating	1.32	16.0	64.8
Canoeing	1.39	16.9	68.3
Fishing	1.28	15.6	63.0
Rowing	1.25	15.2	61.3
Kayaking	1.35	16.5	66.5
Wading	1.32	16.0	64.8

Table 2 Mean number of *E. coli*-borne antibiotic resistance genes ingested by different water users in bathing waters in different categories of water quality: 1) *E. coli* density ≤250 cfu 100 mL^{-1} 2) *E. coli* density 251-500 cfu 100 mL^{-1} 3) *E. coli* density >500 cfu 100 mL^{-1}. *High-contact water sports include those with a high chance of head immersion in the water, low-contact water sports include those activities with a low chance of head immersion.

Discussion

For the first time, a targeted metagenomic method to estimate the abundance and diversity of antibiotic resistance determinants in an indicator bacterium, *E. coli*, is described. Data generated on the relative abundance and diversity of antibiotic resistance genes (ARGs) were used to demonstrate that this approach can facilitate estimations of human exposure to *E. coli*-borne ARGs in coastal bathing waters. Based on the abundance of ARGs per *E. coli* isolated from 13 designated bathing waters in England, it was estimated that the average *E. coli* in this environmental compartment harboured 1.24 ARGs, and ingestion of *E. coli*-borne resistance genes is a certainty when taking part in water sports (100% of water sports sessions in 2016 involved the ingestion of *E. coli* harbouring ARGs). If the threshold for exposure to resistance is raised to ingestion of *E. coli* harbouring 100 or more ARGs, 2.5 million exposure events were estimated to have occurred. This demonstrates that bathing in coastal waters could be a significant means by which a large number of people are exposed to high numbers of genes conferring a wide variety of resistance to antibiotics, including genes borne on mobile genetic elements, harboured by bacteria capable of colonising the human gut (Leonard, et al. 2015). This is likely to be an underestimate of exposure, because genes thought to confer multidrug resistance were excluded from the analyses and mutation based resistance was not considered. These estimates of exposure are greater than those previously estimated by Leonard et al. (2015), when exposure to *E. coli* phenotypically resistant to one group of antibiotics (third-generation cephalosporins) was estimated among water users.

The structured antibiotic resistance gene database (SARG) used in the online analysis pipeline (ARGs-OAP) enabled the classification of antibiotic resistance genes into the types of antibiotic to which identified genes conferred resistance. Genes conferring resistance to beta-lactam antibiotics were most commonly found among both *E. coli* and non-*E. coli* coliform metagenomes, accounting for 20% and 31% respectively of the resistance genes identified. It was also possible to quantify the diversity of resistance gene subtypes in each sample. The highest diversity of ARGs was found among *E. coli* from sample number 2, and diversity was also relatively high in samples 1 and 3. Diversity was otherwise quite consistent among *E. coli* across all sites. A possible explanation for variations in

ARG diversity across sites could be differences in the sources of pollution affecting each site. For example, sites heavily impacted by human activities and those with multiple pollution sources might have a higher abundance and diversity of ARGs compared to more pristine sites (Li, et al. 2015). The approach described has the potential to address important research questions regarding the ecology of antimicrobial resistance determinants in the environment, such as understanding the landscape-scale factors driving high levels of resistance in environments affected by faecal pollution, and to monitor the impact of antibiotic stewardship programmes in the community on ARGs in wastewater and receiving waters. Furthermore, the method provides data on ARGs borne by viable cells and need not be limited to investigating resistance genes in *E. coli*: any culturable bacterium of interest could be used to assess the abundance and diversity of ARGs they harbour. A limitation to this method is that it cannot be used for cells that cannot be cultured in the laboratory, or for samples with very low cell densities, since a minimum amount of genetic material is required for the construction of sequence libraries and for the detection of rare sequences (Tringe and Rubin 2005, Wang, et al. 2013).

On average, a sequencing depth equivalent to 71 genomes was achieved for the samples (Table S1), suggesting the coverage was high enough to sequence a representative from each colony in the majority of samples. In composite samples with large numbers of colonies some colonies may be underrepresented, however a similar coverage of *E. coli* genomes is achieved in each water sample independent of *E. coli* density in the original water samples. Interestingly there was no relationship between *E. coli* number in each pool and the number of annotated resistance genes, suggesting variation in resistance gene relative abundance and diversity is not a function of sampling effort. Compared to conventional metagenomics, where a small proportion of thousands of species and strains are sequenced giving very low coverage of total genetic diversity, this targeted approach not only links antibiotic resistance gene diversity with a specific host species, but also gives relatively high coverage of *E.* coli diversity with the majority of isolates being sequenced. Therefore, it offers a pragmatic, cost effective approach to surveying resistance genes in environmental *E. coli*.

E. coli is a highly useful indicator organism for the purposes of studying ARGs in a variety of matrices because of its widespread use as a faecal indicator bacterium for monitoring the bacteriological quality of aquatic environments, food and drinking water (Odonkor and Ampofo 2013). In addition, data on the abundance and types of ARGs present in viable bacterial cells that are capable of colonising the human gut are provided. Therefore, this approach can be used to understand the abundance and diversity of ARGs harboured by *E. coli* isolated in a range of contexts.

Conclusions

A metagenomic method targeting *E. coli* and other coliforms was used to investigate the abundance and diversity of antibiotic resistance genes in a commonly measured indicator bacterium, *E. coli,* and facilitated an estimation of human exposure to ARGs borne by *E. coli* in coastal bathing waters. On average, *E. coli* were found to harbour 1.24 ARGs per cell. In 2016 all water sports sessions in England were estimated to result in ingestion of one or more *E. coli*-borne ARGs and 2.5 million sessions involved the ingestion of *E. coli* harbouring 100 or more ARGs. This targeted metagenomic approach has a variety of possible applications to facilitate the monitoring of resistance genes, as well as investigations into of the ecology of antibiotic resistance genes in various environments and matrices.

Funding

This work was supported by the University of Exeter and the Chinese University of Hong Kong.

Acknowledgements

Thanks to Dr Jonathan Porter and the staff at the Environment Agency for supplying us with *E. coli* specimens for this work. Thanks also to the Exeter Sequence Service facility. The Exeter Sequencing

Service is supported by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), the Wellcome Trust Multi User Equipment Award (WT101650MA), the Medical Research Council Clinical Infrastructure Funding (MR/M008924/1) and the Biotechnology and Biological Sciences Research Council Longer and Larger grants award (BB/K003240/1). The authors have no conflicts of interest to declare.

References

Ashbolt NJ, Amezquita A, Backhaus T *et al.* Human Health Risk Assessment (HHRA) for environmental development and transfer of antibiotic resistance. *Environmental Health Perspectives* 2013;**121**: 993-1001.

Blackburn H, O'Neill R, Rangeley-Wilson C *et al.* Flushed away: How sewage is still polluting the rivers of England and Wales: World Wildlife Fund, 2017.

Costelloe C, Williams OM, Montgomery AA *et al.* Antibiotic Prescribing in Primary Care and Antimicrobial Resistance in Patients Admitted to Hospital with Urinary Tract Infection: A Controlled Observational Pilot Study. *Antibiotics (Basel)* 2014;**3**: 29-38.

Environment Agency. The Microbiology of Drinking Water (2009) - Part 4 - Methods for the isolation and enumeration of coliform bacteria and Escherichia coli (including O157:H7)*Standing Committee of Analysts blue books: Methods for the Examination of Waters and Associated Materials* 2009.

Environment Agency. Find a bathing water volume 2017: Environment Agency, 2017.

Ercolini D. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Appl Environ Microbiol* 2013;**79**: 3148-55.

European Parliament Council of the European Union. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC. In: Union EPatCotE (ed.), 2006.

Handel N, Otte S, Jonker M *et al.* Factors That Affect Transfer of the IncI1 beta-Lactam Resistance Plasmid pESBL-283 between E-coli Strains. *Plos One* 2015;**10**.

Ihekweazu C, Barlow M, Roberts S *et al.* Outbreak of E. coli O157 infection in the south west of the UK: risks from streams crossing seaside beaches. *Euro surveillance : bulletin europeen sur les maladies transmissibles = European communicable disease bulletin* 2006;**11**: 128-30.

International Organization for Standardization. ISO/TS 16649-3:2005 Microbiology of food and animal feeding stufffs -- Horizontal method for the enumeration of beta-glucuronidase-positive Escherichia coli *Part 3: Most probable number technique using 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide*, 2005.

International Organization for Standardization. Water quality -- Enumeration of Escherichia coli and coliform bacteria -- Part 1: Membrane filtration method for waters with low bacterial background flora*Water quality, Microbiological Methods*, 2014a.

International Organization for Standardization. Water quality -- Requirements for the comparison of the relative recovery of microorganisms by two quantitative methods*Water quality, Microbiology*, 2014b.

Johnson JR, Johnston B, Clabots C *et al.* Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. *Clinical Infectious Diseases* 2010;**51**: 286-94.

Kantar TNS. The Great Britain Day Visitor 2016 annual report, 2017.

Leonard A, Zhang L, Balfour A *et al.* Human recreational exposure to antibiotic resistant bacteria in coastal bathing waters. *Environment International* 2015;**82**: 92-100.

Levy SB, FitzGerald GB, Macone AB. Spread of antibiotic-resistant plasmids from chicken to chicken and from chicken to man. *Nature* 1976;**260**: 40-2.

Li B, Yang Y, Ma L *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME journal* 2015;**9**: 2490-502.

Logares R, Haverkamp TH, Kumar S *et al.* Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods* 2012;**91**: 106-13.

Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced Escherichia coli genomes. *Microb Ecol* 2010;**60**: 708-20.
Monaghan J, Hutchinson M. Monitoring microbial food safety of fresh produce. In: Food Standards Agency, Horitcultural Development Company (eds.): Horiticultural Development Company,, 2010.
Odonkor ST, Ampofo JK. Escherichia coli as an indicator of bacteriological quality of water: an overview. *Microbiology Research* 2013;**4**: 5-11.
Office for National Statistics. Population estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2016. ONS, 2016.
Pruden A, Pei R, Storteboom H *et al.* Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environ Sci Technol* 2006;**40**: 7445-50.
Review on antimicrobial resistance. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations, 2014.
Review on antimicrobial resistance. Antimicrobials in agriculture and the environment: reducing unnecessary use and waste. In: O'Neill J (ed.) volume 2016, 2015.
Russo TA, Johnson JR. Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. *Microbes and Infection* 2003;**5**: 449-56.
Shuval H. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *Journal of Water & Health* 2003;**1**: 53-64.
Soraas A, Sundsfjord A, Sandven I *et al.* Risk Factors for Community-Acquired Urinary Tract Infections Caused by ESBL-Producing Enterobacteriaceae -A Case-Control Study in a Low Prevalence Country. *PLoS One* 2013;**8**: e69581.
Tringe SG, Rubin EM. Metagenomics: DNA sequencing of environmental samples. *Nature reviews Genetics* 2005;**6**: 805-14.
Wang J, McLenachan PA, Biggs PJ *et al.* Environmental bio-monitoring with high-throughput sequencing. *Brief Bioinform* 2013;**14**: 575-88.
Yang Y, Jiang X, Chai B *et al.* ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. *Bioinformatics* 2016;**32**: 2346-51.